

# Weakly Secure MDS Codes for Simple Multiple Access Networks

Son Hoang Dau\*, Wentu Song<sup>†</sup>, Chau Yuen<sup>†</sup>

Singapore University of Technology and Design, Singapore

Emails: {<sup>\*</sup>*sonhoang\_dau*, <sup>†</sup>*wentu\_song*, <sup>‡</sup>*yuenchau*}@sutd.edu.sg

*Abstract*—We consider a simple multiple access network (SMAN), where  $k$  sources of unit rates transmit their data to a common sink via  $n$  relays. Each relay is connected to the sink and to certain sources. A coding scheme (for the relays) is *weakly secure* if a passive adversary who eavesdrops on less than  $k$  relay-sink links cannot reconstruct the data from each source.

We show that there exists a weakly secure maximum distance separable (MDS) coding scheme for the relays if and only if every subset of  $\ell$  relays must be collectively connected to at least  $\ell + 1$  sources, for all  $0 < \ell < k$ . Moreover, we prove that this condition can be verified in polynomial time in  $n$  and  $k$ . Finally, given a SMAN satisfying the aforementioned condition, we provide another polynomial time algorithm to trim the network until it has a sparsest set of source-relay links that still supports a weakly secure MDS coding scheme.

## I. INTRODUCTION

A simple multiple access network (SMAN) is a two-hop network, where some  $k$  independent sources transmit their data to a common sink via  $n$  relays. We use  $(n, k)$ -SMAN to refer to such network. An example of a  $(6, 4)$ -SMAN is illustrated in Fig. 1. Simple multiple access networks were studied in the recent work of Yao *et al.* [1] (to model the problem of decentralized distribution of keys from a pool among the wireless nodes), Halbawi *et al.* [2], and Dau *et al.* [3], [4], [5]. The model of SMAN considered in [2] is more general in the sense that the sources are assumed to have arbitrary rates. However, it was shown in [4], [5] that as far as the problem of constructing error-correcting codes for the relays is concerned, considering unit-rate sources is sufficient. Interestingly, the code design problem for SMAN was also shown in [4], [5] to be equivalent to the code design problem for weakly secure cooperative data exchange [6], [7].

Error correction for the general multiple access network was first investigated in the work of Dikaliotis *et al.* [8]. The coding schemes derived in [8] are packetized over large fields, which are of sizes at least exponential in the number of sources. While SMAN is a special case of multiple access network [8], the authors of [2], [4], [5] focused more on designing error-correcting codes over small fields, whose sizes are linear in  $n$  and  $k$ . Various new problems on balance and sparsity of the network were also investigated in [3], [5].

In this paper we study the security aspect of the coding schemes used for the relays in an  $(n, k)$ -SMAN. More specifically, we focus on the *weak security* of such coding schemes against a *passive adversary*, which eavesdrops on the relay-sink links. Suppose that each source transmits a

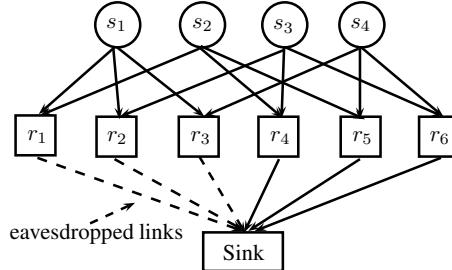


Fig. 1: An example of a (6, 4)-SMAN. Three relay-sink links (dashed) are eavesdropped. The question is: can we prevent the adversary from learning about each individual source packet?

single packet, which is an element of some finite field  $\mathbb{F}_q$ , to the sink. All source packets are assumed to be independent and randomly distributed over  $\mathbb{F}_q$ . The coding scheme for the relays is *weakly secure* if an adversary that eavesdrops on at most  $k - 1$  relay-sink links gains *no information* (in Shannon's sense) about each particular source packet. In the context of decentralized key distribution [1], a wireless node (corresponding to the sink in the SMAN) contacts its neighbors (corresponding to the relays in the SMAN) to retrieve  $k$  secret keys  $s_i \in \mathbb{F}_q$  ( $1 \leq i \leq k$ ). Each of its neighbors possesses some of these  $k$  keys and transmits one (coded) packet in  $\mathbb{F}_q$  to that node. In that scenario, a weakly secure coding scheme for the corresponding SMAN would guarantee that an adversary that eavesdrops on at most  $k - 1$  transmissions cannot determine explicitly any secret key. Note that Yao *et al.* [1] only considered an active adversary who can corrupt the transmissions. In this work, we assume the presence of both an active adversary and a passive adversary. Note that these two adversaries may be independent of each other. In other words, they may attack different sets of links.

The concept of weak security was first discussed by Yamamoto [9] in the context of ramp secret sharing scheme. After Yamamoto [9], weak security was also discovered by Bhattacharjee and Narayanan [10] in a more general context of network coding. Weak security is important in practice since it guarantees that no meaningful information is leaked to the adversary, and often requires no additional overhead. For example, suppose that the adversary obtains the coded packet  $x_1 + x_2$  where  $x_1$  and  $x_2$  are packets from the sources  $s_1$  and  $s_2$ , respectively. Then the adversary would not be able to determine either  $x_1$  or  $x_2$ , as from its point of view, both  $x_1$  and  $x_2$  are completely random variables.

In this work we limit ourselves to maximum distance separable (MDS) coding schemes (see Section II for definition). Our main contributions are summarized below.

- We establish a necessary and sufficient condition for the existence of a weakly secure MDS coding scheme for the relays. More specifically, there exists a weakly secure MDS coding scheme for the relays if and only if every subset of  $\ell$  relays must be collectively connected to at least  $\ell + 1$  sources, for all  $0 < \ell < k$ . Moreover, this condition, referred to as the *Weak Security Condition*, can be verified in polynomial time in  $n$  and  $k$ .
- Given a SMAN satisfying the Weak Security Condition, we provide a polynomial time algorithm to trim the network by removing certain source-relay links until it has the sparsest set of source-relay links that still supports a weakly secure MDS coding scheme. This algorithm is similar to the algorithm used to find a maximum matching in a bipartite graph that deletes edges of the graph one by one until all remaining edges form a matching.
- We also study the so-called *block security*, which is a generalization of weak security, and characterize the block security level of an arbitrary SMAN.

The first conclusion above describes the additional requirement on the source-relay links if a passive adversary is also present. Indeed, an MDS code implemented at the relays allows the sink to tolerate a maximum number of  $\lfloor (n - k + 1)/2 \rfloor$  corrupted relay/links. Such an MDS code exists if and only if the SMAN satisfies the MDS Condition [3], [4], [5], [7]: every subset of  $\ell$  relays must be collectively connected to at least  $\ell$  sources, for all  $\ell \leq k$ . Comparing the MDS Condition and the Weak Security Condition, we conclude that more source-relay links are required to defend both an active and a passive adversary. Hence, a SMAN that survives the most powerful active adversary, which can corrupt  $\lfloor (n - k + 1)/2 \rfloor$  relays/links, may not be weakly secure against a passive adversary.

The paper is organized as follows. Necessary notation and definitions are provided in Section II. The weak security for SMAN is discussed in Section III. The extension of weak security to block security is investigated in Section IV.

## II. PRELIMINARIES

Let  $\mathbb{F}_q$  denote the finite field with  $q$  elements. Let  $[n]$  denote the set  $\{1, 2, \dots, n\}$ . For a  $k \times n$  matrix  $\mathbf{M}$ , for  $i \in [k]$  and  $j \in [n]$ , let  $\mathbf{M}_i$  and  $\mathbf{M}[j]$  denote the row  $i$  and the column  $j$  of  $\mathbf{M}$ , respectively. We define below standard notions from coding theory (for instance, see [11]).

The *support* of a vector  $\mathbf{u} = (u_1, \dots, u_n) \in \mathbb{F}_q^n$  is the set  $\text{supp}(\mathbf{u}) = \{i \in [n] : u_i \neq 0\}$ . The (Hamming) distance between two vectors  $\mathbf{u}$  and  $\mathbf{v}$  of  $\mathbb{F}_q^n$  is defined to be  $d(\mathbf{u}, \mathbf{v}) = |\{i \in [n] : u_i \neq v_i\}|$ . A  $k$ -dimensional subspace  $\mathcal{C}$  of  $\mathbb{F}_q^n$  is called a linear  $[n, k, d]_q$  (*error-correcting*) code over  $\mathbb{F}_q$  if the minimum distance  $d(\mathcal{C})$  between any pair of distinct vectors in  $\mathcal{C}$  is equal to  $d$ . Sometimes we may use the notation  $[n, k]_q$  or just  $[n, k]$  for the sake of simplicity. The vectors in  $\mathcal{C}$  are called *codewords*. A *generator matrix*  $\mathbf{G}$  of an  $[n, k]_q$  code

$\mathcal{C}$  is a  $k \times n$  matrix whose rows are linearly independent codewords of  $\mathcal{C}$ . Then  $\mathcal{C} = \{\mathbf{x}\mathbf{G} : \mathbf{x} \in \mathbb{F}_q^k\}$ . The well-known Singleton bound ([11, Ch. 1]) states that for any  $[n, k, d]_q$  code, it holds that  $d \leq n - k + 1$ . If the equality is attained, the code is called *maximum distance separable* (MDS).

**Definition 1.** An  $(n, k)$  simple multiple access network  $((n, k)\text{-SMAN}$  for short) is a network that consists of

- $k$  independent sources  $s_1, \dots, s_k$  of unit rates, one sink, and  $n$  relays  $r_1, \dots, r_n$ , where  $n \geq k$ , and
- some directed edges of capacity one that connect certain source-relay pairs and one directed edge of capacity one that connects each relay to the sink.

An  $(n, k)\text{-SMAN}$  can be represented by an *adjacency matrix*  $\mathbf{M} = (m_{i,j}) \in \mathbb{F}_2^{k \times n}$  where  $m_{i,j} = 1$  if and only if the source  $s_i$  is connected to the relay  $r_j$ .

Let  $\mathbf{X} = (X_1, \dots, X_k)$  be a vector of independent and identically uniformly distributed random variables over  $\mathbb{F}_q$ . We assume that the vector of source packets is  $\mathbf{x} = (x_1, \dots, x_k)$ , a realization of  $\mathbf{X}$ . A linear coding scheme for an  $(n, k)\text{-SMAN}$  is represented by a  $k \times n$  matrix  $\mathbf{G} = (g_{i,j})$  over  $\mathbb{F}_q$ . The coding rule for the relays is as follows:  $r_j$  ( $j \in [n]$ ) creates and transmits the coded packet  $\mathbf{x}\mathbf{G}[j]$  to the sink. We refer to  $\mathbf{G}$  as the *encoding matrix* of the coding scheme. Note that  $g_{i,j}$  must be zero whenever  $m_{i,j} = 0$ . If  $\mathbf{G}$  generates a linear code that can correct  $t$  errors then the sink can still determine all  $k$  source packets under the presence of at most  $t$  erroneous coded packets sent from some  $t$  relays.

A coding scheme based on  $\mathbf{G}$  is *weakly secure* if the conditional entropy

$$H(X_i | \{\mathbf{x}\mathbf{G}[j] : j \in E\}) = H(X_i),$$

for every  $i \in [k]$  and for every subset  $\emptyset \neq E \subset [n]$ ,  $|E| < k$ . In words, a coding scheme is *weakly secure* if an adversary that eavesdrops on at most  $k - 1$  coded packets transmitted on different relay-sink links obtains no information about each particular source packet. Note that we always assume that  $\text{rank}_q(\mathbf{G}) = k$ . Hence, obviously an adversary that eavesdrops on certain  $k$  linearly independent coded packets can always retrieve all  $k$  source packets.

## III. WEAK SECURITY FOR SMAN

### A. Necessary and Sufficient Condition for Weak Security

We first derive a necessary and sufficient condition on the links between sources and relays for a SMAN to support a weakly secure MDS coding scheme.

**Theorem 1.** An  $(n, k)\text{-SMAN}$  supports a weakly secure MDS coding scheme, i.e. there exists a weakly secure MDS coding scheme for the relays over some finite field  $\mathbb{F}_q$ , if and only if every subset of  $\ell$  relays must be collectively connected to at least  $\ell + 1$  sources, for all  $0 < \ell < k$ . In other words, it requires that

$$|\cup_{j \in J} \text{supp}(\mathbf{M}[j])| \geq |J| + 1, \quad \forall \emptyset \neq J \subset [n], \quad |J| < k, \quad (1)$$

where  $\mathbf{M}[j]$  is the  $j$ th column of the adjacency matrix  $\mathbf{M}$ . We refer to (1) as the *Weak Security Condition for SMAN*.

We need a few lemmas for the proof of Theorem 1.

**Lemma 1** ([12]). *The  $k \times n$  matrix  $\mathbf{G}$  is a generator matrix of an  $[n, k, d]_q$  error-correcting code if and only if every  $n-d+1$  columns of  $\mathbf{G}$  has rank  $k$ .*

**Lemma 2.** *A coding scheme based on the matrix  $\mathbf{G}$  for an  $(n, k)$ -SMAN is weakly secure if and only if every  $k-1$  columns of  $\mathbf{G}$  generates an error-correcting code of minimum distance at least two.*

*Proof:* This is a corollary of [13, Lemma 3]. More details can be found in Appendix A. ■

**Lemma 3.** *If the  $\ell \times k$  matrix  $\mathbf{A}$  generates a  $[k, \ell, d \geq 2]_q$  error-correcting code then*

$$|\cup_{j \in J} \text{supp}(\mathbf{A}_j)| \geq |J| + 1, \quad \forall \emptyset \neq J \subseteq [\ell]. \quad (2)$$

*Proof:* Suppose that  $\mathbf{A}$  generates a code of minimum distance at least two but (2) is violated. Then there exists  $\emptyset \neq J \subseteq [\ell]$  such that

$$|\cup_{j \in J} \text{supp}(\mathbf{A}_j)| \leq |J|. \quad (3)$$

We aim to obtain a contradiction.

Let  $I \subseteq [k] \setminus \cup_{j \in J} \text{supp}(\mathbf{A}_j)$  such that  $|I| = k - |J|$ . Moreover, let  $L \subset [k]$  such that  $L \supseteq I$  and  $|L| = k - 1$ . Let  $\mathbf{A}[L]$  be the  $\ell \times (k-1)$  submatrix of  $\mathbf{A}$  that consists of columns of  $\mathbf{A}$  indexed by the elements in  $L$ . Then according to Lemma 1, we have

$$\text{rank}_q(\mathbf{A}[L]) = \ell. \quad (4)$$

On the other hand, we claim that the  $|J|$  rows of  $\mathbf{A}[L]$  indexed by the elements in  $J$  has rank at most  $|J|-1$ . As the remaining  $\ell - |J|$  rows of  $\mathbf{A}[L]$  has rank at most  $\ell - |J|$ , we deduce that

$$\text{rank}_q(\mathbf{A}[L]) \leq (|J|-1) + (\ell - |J|) < \ell. \quad (5)$$

From (4) and (5) we obtain a contradiction.

We now prove that our aforementioned claim is correct. Consider the submatrix  $\mathbf{A}_J[L]$  that consists of rows of  $\mathbf{A}[L]$  indexed by the elements of  $J$ . Due to (3) and our assumption that  $L \supseteq I$ , the submatrix  $\mathbf{A}_J[L]$  has at least  $k - |J|$  all-zero columns. Since  $|L| = k - 1$ ,  $\mathbf{A}_J[L]$  has  $k - 1$  columns. Therefore, it has at most  $|J| - 1$  nonzero columns. Hence,  $\text{rank}_q(\mathbf{A}_J[L]) \leq |J| - 1$ , as claimed. ■

**Remark 1.** The result in Lemma 3 can be extended to  $d \geq d'$  for any  $d' \geq 1$  by replacing  $|J| + 1$  with  $|J| + d' - 1$  in (2).

**Lemma 4.** *Let  $\mathbf{P}$  be a  $(k-1) \times k$  0-1 matrix. Let  $\text{var}(\mathbf{P})$  be the matrix obtained from  $\mathbf{P}$  by replacing every nonzero entry of  $\mathbf{P}$  by some indeterminate  $\xi_{i,j}$  over  $\mathbb{F}_q$ . Suppose that all of these indeterminates are independent. Let  $f(\text{var}(\mathbf{P})) = \prod_Q \det(Q)$ , where the product is taken over all  $k$  submatrices  $Q$  of order  $k-1$  of  $\text{var}(\mathbf{P})$ . Then  $f(\text{var}(\mathbf{P}))$ , which is a multivariable polynomial in  $\mathbb{F}_q[\dots, \xi_{i,j}, \dots]$ , is not identically zero if and only if*

$$|\cup_{j \in J} \text{supp}(\mathbf{P}_j)| \geq |J| + 1, \quad \forall \emptyset \neq J \subseteq [k-1]. \quad (6)$$

*Proof:* The proof follows from [3, Lemma 2-4]. More details can be found in Appendix B. ■

We are now in position to prove Theorem 1.

*Proof of Theorem 1:*

**Only-If.** Suppose that there exists a weakly secure MDS coding scheme for an  $(n, k)$ -SMAN described by the adjacency matrix  $\mathbf{M}$ . We aim to prove that the Weak Security Condition (1) holds. Let  $\mathbf{G}$  be the encoding matrix of the weakly secure MDS coding scheme. Note that as  $\mathbf{G}$  generates an MDS code, every subset of  $k-1$  columns of  $\mathbf{G}$  is always linearly independent [11, Ch. 11]. Hence, by Lemma 2, every set of  $k-1$  columns of  $\mathbf{G}$  must generate a  $[k, k-1, 2]$  error-correcting code. Note here that  $\text{supp}(\mathbf{G}[j]) \subseteq \text{supp}(\mathbf{M}[j])$  for all  $j \in [n]$ . Hence, by applying Lemma 3 to all  $(k-1) \times k$  matrices corresponding to all subsets of  $k-1$  columns of  $\mathbf{G}$ , it is straightforward that the Weak Security Condition holds.

**If.** We assume that the Weak Security Condition holds, i.e.

$$|\cup_{j \in J} \text{supp}(\mathbf{M}[j])| \geq |J| + 1, \quad \forall \emptyset \neq J \subseteq [n], |J| \leq k-1.$$

We aim to show that there exists a weakly secure MDS coding scheme for the corresponding  $(n, k)$ -SMAN.

Using the same notation as in Lemma 4, let  $\text{var}(\mathbf{M}) = (v_{i,j})$  where  $v_{i,j} = 0$  if  $m_{i,j} = 0$  and  $v_{i,j} = \xi_{i,j}$  if  $m_{i,j} \neq 0$ . Here  $\xi_{i,j}$ 's are independent indeterminates. For each submatrix  $\mathbf{P}'$  of size  $k \times (k-1)$  of  $\mathbf{M}$ , let  $\mathbf{P}$  be its transpose and  $\text{var}(\mathbf{P})$  the corresponding (transposed) submatrix of  $\text{var}(\mathbf{M})$ . We henceforth refer to such a matrix  $\mathbf{P}$  as a *transposed submatrix* of  $\mathbf{M}$ . Note that the Weak Security Condition (1) on  $\mathbf{M}$  implies the condition (6) on every transposed submatrix  $\mathbf{P}$  of size  $(k-1) \times k$  of  $\mathbf{M}$ . Hence, by Lemma 4, the polynomial  $f(\text{var}(\mathbf{P}))$  is not identically zero. Let

$$F(\text{var}(\mathbf{M})) = \prod_{\mathbf{P}} f(\text{var}(\mathbf{P})) \in \mathbb{F}_q[\dots, \xi_{i,j}, \dots],$$

where the product is taken over all transposed submatrices  $\mathbf{P}$  of size  $(k-1) \times k$  of  $\mathbf{M}$ . Then  $F(\text{var}(\mathbf{M})) \neq 0$ .

It is obvious that the Weak Security Condition (1) implies the MDS Condition [4], [5], [7], which requires that every subset of  $\ell$  relays must be collectively connected to at least  $\ell$  sources, for all  $\ell \leq k$ . Hence, if  $f(\text{var}(\mathbf{M}))$  is the product of determinants of all submatrices of order  $k$  of  $\text{var}(\mathbf{M})$  then  $f(\text{var}(\mathbf{M})) \neq 0$ , according to [3, Lemma 2-4]. Therefore

$$F^{\text{ext}}(\text{var}(\mathbf{M})) \triangleq f(\text{var}(\mathbf{M})) \times F(\text{var}(\mathbf{M})) \neq 0.$$

Hence, according to [14, Lemma 4], for sufficiently large  $q$ , there exists  $g_{i,j} \in \mathbb{F}_q$  (for  $(i, j)$  where  $m_{i,j} = 1$ ) such that

$$F^{\text{ext}}(\text{var}(\mathbf{M}))(\dots, g_{i,j}, \dots) \neq 0.$$

As a consequence,

$$f(\text{var}(\mathbf{P}))(\dots, g_{i,j}, \dots) \neq 0, \quad (7)$$

for every transposed submatrix  $\mathbf{P}$  of size  $(k-1) \times k$  of  $\mathbf{M}$ . Let  $\mathbf{G} = (g_{i,j})$  (if  $m_{i,j} = 0$  we set  $g_{i,j} = 0$ ). Then thanks to (7), every transposed submatrix  $\mathbf{A}$  of size  $(k-1) \times k$  of  $\mathbf{G}$  satisfies the following property: all submatrices of order  $k-1$  of  $\mathbf{A}$  are invertible. Hence, according to [11, Ch. 11], every set of  $k-1$  columns of  $\mathbf{G}$  generates an MDS  $[k, k-1, 2]_q$  error-correcting code. Thus, by Lemma 2, the coding scheme based on  $\mathbf{G}$  is

weakly secure. Moreover, as  $f(\text{var}(\mathbf{M}))(\dots, g_{i,j}, \dots) \neq 0$  as well, it follows that every submatrix of order  $k$  of  $\mathbf{G}$  is invertible. Thus,  $\mathbf{G}$  also generates an MDS code. ■

**Remark 2.** Theorem 1 shows what the additional cost is (in terms of source-relay links) when a passive adversary is also present, on top of an active adversary. More specifically, while defending against an active adversary requires that every subset of  $\ell$  relays must be collectively connected to at least  $\ell$  sources, for all  $\ell \leq k$ , defending against both adversaries requires that every subset of  $\ell$  relays must be collectively connected to at least  $\ell + 1$  sources, for all  $0 < \ell < k$ .

### B. Verification of Weak Security Condition in Polynomial Time

While designing a weakly secure MDS coding scheme for a given  $(n, k)$ -SMAN may require non-polynomial time (as random coding over finite fields with exponentially large sizes is used), verifying whether a SMAN supports a weakly secure MDS coding scheme can be done in polynomial time. We prove this fact below using a proper modification of the proof of [5, Lemma 10]. We first present a simple lemma. Its proof is similar to the proof of [3, Lemma 4] and can be found in Appendix C.

**Lemma 5.** *The Weak Security Condition (1) is equivalent to the following:*

$$|\cup_{i \in I} \text{supp}(\mathbf{M}_i)| \geq n - k + |I| + 1, \quad \forall \emptyset \neq I \subsetneq [k]. \quad (8)$$

**Proposition 1.** *The Weak Security Condition (1) can be verified in polynomial time in  $n$  and  $k$ .*

*Proof:* By Lemma 5, it suffices to prove that (8) can be verified in polynomial time for all  $\emptyset \neq I \subseteq [k] \setminus \{i_0\}$ , for every  $i_0 \in [k]$ . Without loss of generality, let  $i_0 = k$ . The other cases can be proved in the same manner. We associate with  $\mathbf{M}$  a network  $\mathcal{N}^k(\mathbf{M})$  constructed as follows. The set of nodes of  $\mathcal{N}^k(\mathbf{M})$  consists of

- a source node  $s$ ,
- $n$  packet nodes  $s_1, \dots, s_n$ ,
- $k - 1$  coding nodes  $r_1, \dots, r_{k-1}$ ,
- $k - 1$  broadcast nodes  $b_1, \dots, b_{k-1}$ ,
- $k - 1$  sink nodes  $t_1, \dots, t_{k-1}$ .

To simplify the notation, set  $R_i = \text{supp}(\mathbf{M}_i)$ . The set of directed edges of  $\mathcal{N}^k(\mathbf{M})$  consists of

- one edge of capacity *one* from  $s$  to  $s_i$ ,  $\forall i \in [n]$ ,
- one edge of capacity *infinity* from  $s_j$  to  $r_i$  if  $j \in R_i$ ,
- one edge of capacity *one* from  $r_i$  to  $b_i$ ,  $\forall i \in [k-1]$ ,
- one edge of capacity *infinity* from  $r_i$  to  $t_i$ ,  $\forall i \in [k-1]$ ,
- one edge of capacity *infinity* from  $b_i$  to  $t_j$ ,  $\forall i, j \in [k-1]$ .

For instance, for the  $(6, 4)$ -SMAN in Fig. 1, the corresponding adjacency matrix is

$$\mathbf{M} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}, \quad (9)$$

and the corresponding network  $\mathcal{N}^4(\mathbf{M})$  is depicted in Fig. 2.

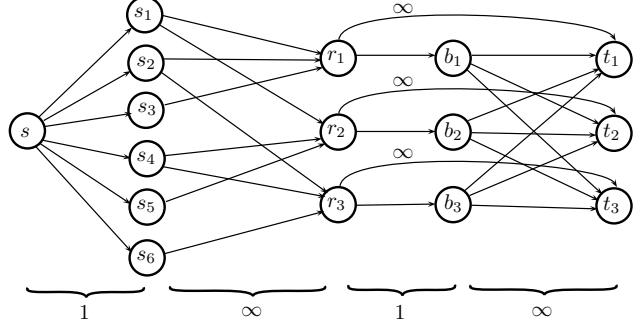


Fig. 2: The network  $\mathcal{N}^4(\mathbf{M})$  associated to  $\mathbf{M}$  in (9) when  $i_0 = k = 4$ .

A cut  $(S, T)$  of a network is a partition of the set of nodes of that network into two parts, namely  $S$  and  $T$ . We are only interested in cuts that *separate* the source and some sink, i.e.  $S$  contains the source and  $T$  contains some sink. Let  $c(u, v)$  denote the capacity of an edge  $(u, v)$  in a network. Then the capacity of a cut  $(S, T)$  is defined as

$$c(S, T) = \sum_{u \in S, v \in T} c(u, v).$$

Consider the following *Min-Cut* Condition for  $\mathcal{N}^k(\mathbf{M})$ : the capacity of every cut that separates  $s$  and any sink is at least  $n$ . According to the Network Flow Algorithm (Ford-Fulkerson Algorithm), we can verify the Min-Cut Condition for  $\mathcal{N}^k(\mathbf{M})$  in polynomial time. Therefore, it suffices for our purpose to show that the condition (8) restricted to those  $I \subseteq [k-1]$  (for  $\mathbf{M}$ ) is equivalent to the Min-Cut Condition (for  $\mathcal{N}^k(\mathbf{M})$ ).

Suppose that the Min-Cut Condition for  $\mathcal{N}^k(\mathbf{M})$  holds. We aim to prove that (8) restricted to those  $I \subseteq [k-1]$  also holds for  $\mathbf{M}$ . Let  $I$  be an arbitrary nonempty subset of  $[k-1]$ . Recall that we use  $R_i$  to denote  $\text{supp}(\mathbf{M}_i)$ . Consider a cut  $(S, T)$  where

$$T = \{t_i : i \in I\} \cup \{b_i : 1 \leq i \leq k-1\} \cup \{r_i : i \in I\} \\ \cup \{s_j : j \in \cup_{i \in I} R_i\}.$$

Then the capacity of  $(S, T)$  is

$$c(S, T) = \sum_{j \in \cup_{i \in I} R_i} c(s, s_j) + \sum_{i \notin I} c(r_i, b_i) = |\cup_{i \in I} R_i| + k - 1 - |I|.$$

As  $c(S, T) \geq n$ , we have

$$|\cup_{i \in I} R_i| \geq n - k + |I| + 1.$$

Conversely, suppose that (8) restricted to those  $I \subseteq [k-1]$  holds. We need to prove that  $c(S, T) \geq n$  for every cut  $(S, T)$  that separates  $s$  and some sink. Suppose that  $\{t_i\}_{i \in I'} \subseteq T$ , where  $\emptyset \neq I' \subseteq [k-1]$ , and that  $t_i \notin T$  if  $i \notin I'$ . If  $c(S, T) = \infty$  then it is larger than  $n$  trivially. Now suppose that  $c(S, T) < \infty$ . Then  $(S, T)$  does not contain any edge of the form  $(s_j, r_i)$ ,  $(r_i, t_i)$ , or  $(b_j, t_i)$ , as these have capacity infinity. Hence,  $T$  must contain the following nodes

- $t_i$  for all  $i \in I'$ , because of our definition of  $(S, T)$ ,
- $b_j$  for all  $j \in [k-1]$ , as  $c(b_j, t_i) = \infty$  for every  $j$  and  $i$ ,
- $r_i$  for all  $i \in I'$ , as  $c(r_i, t_i) = \infty$  for every  $i$ ,

Let  $I$  be the subset of  $[k - 1]$  that satisfies

$$T \cap \{r_i\}_{i \in [k-1]} = \{r_i\}_{i \in I}.$$

Then  $I' \subseteq I$ . Since  $c(s_j, r_i) = \infty$  when  $j \in R_i$ , the set  $T$  must also contain the packet nodes  $s_j$  if  $j \in R_i$  for some  $i \in I$ . Therefore,

$$\begin{aligned} c(S, T) &\geq \sum_{j \in \cup_{i \in I} R_i} c(s, s_j) + \sum_{i \notin I} c(r_i, b_i) \\ &= |\cup_{i \in I} R_i| + k - 1 - |I| \\ &\geq (n - k + |I| + 1) + k - 1 - |I| = n. \end{aligned}$$

We complete the proof.  $\blacksquare$

### C. Trimming SMAN While Preserving Weak Security in Polynomial Time

Given an  $(n, k)$ -SMAN that satisfies the Weak Security Condition (1), Theorem 2 states that one can trim the network to obtain a sparsest possible network where the Weak Security Condition is still satisfied. Moreover, the trimming process can be done in polynomial time in  $n$  and  $k$ . We prove this theorem in Appendix D. Note that we use here the equivalent statement of the Weak Security Condition stated in Lemma 5.

**Theorem 2.** *For each  $i \in [k]$  let  $R_i$  be an arbitrary subset of  $[n]$  ( $n \geq k$ ). Suppose that*

$$|\cup_{i \in I} R_i| \geq n - k + |I| + 1, \quad \forall \emptyset \neq I \subsetneq [k]. \quad (10)$$

*Then for every  $i \in [k]$  there exists a subset  $R'_i \subseteq R_i$  such that*

- $|\cup_{i \in I} R'_i| \geq n - k + |I| + 1, \forall \emptyset \neq I \subsetneq [k]$ ,
- $|R'_i| = n - k + 2$ , for all  $i \in [k]$ .

*Moreover, such subsets  $R'_i$  can be found in polynomial time.*

## IV. EXTENSION TO BLOCK SECURITY

In this section we extend our result on weak security to a more general concept of *block security* (or *security against guessing* in some other works in the network coding literature).

The coding scheme for an  $(n, k)$ -SMAN based on a  $k \times n$  encoding matrix  $\mathbf{G}$  is  $b_\ell$ -block secure against a passive adversary of strength  $\ell$  ( $\ell < k$ ) if the conditional entropy

$$H(\{X_j : j \in B\} \mid \{\mathbf{X}\mathbf{G}[j] : j \in E\}) = H(\{X_j : j \in B\}),$$

for every subset  $B \subset [k]$ ,  $|B| \leq b_\ell$ , and for every subset  $E \subset [n]$ ,  $|E| \leq \ell$ . In words, a coding scheme is  $b_\ell$ -block secure against an adversary of strength  $\ell$  if an adversary that eavesdrops on at most  $\ell$  relay-sink packets obtains no information about each subset of at most  $b_\ell$  source packets. In that case, even if the adversary can guess correctly some  $b_\ell - 1$  source packets, it still gains no information about any other packet.

**Theorem 3.** *An  $(n, k)$ -SMAN supports an MDS coding scheme that is  $b_\ell$ -block secure against an adversary of strength  $\ell$  if and only if*

$$|\cup_{j \in J} \text{supp}(\mathbf{M}[j])| \geq |J| + b_\ell, \quad \forall \emptyset \neq J \subset [n], \quad |J| \leq \ell, \quad (11)$$

where  $\mathbf{M}[j]$  denotes the  $j$ th column of the adjacency matrix  $\mathbf{M}$ . Note that in the right-hand side of (11), the first term  $|J|$

corresponds to the MDS Condition, while the second term  $b_\ell$  corresponds to the block security level.

Theorem 3 characterizes the block security level of an MDS coding scheme for SMAN based on the density of the source-relay links. In a special case where the SMAN is densest, i.e. each source is connected to all relays, then according to [15], a Cauchy matrix would provide the best level of block security, which is  $b_\ell = k - \ell$ . The proof of Theorem 3 follows the same idea of that of Theorem 1, using a generalized version of Lemma 3 (see Remark 1). We omit the proof.

While the Weak Security Condition can be verified in polynomial time, it is not known whether a similar conclusion holds for block security. More specifically, given an  $(n, k)$ -SMAN and a sequence  $\{b_\ell\}_1^{k-1}$ , whether (11) can be verified in polynomial time is still an open question.

## ACKNOWLEDGMENT

This work is funded by iTrust.

## REFERENCES

- [1] H. Yao, T. Ho, and C. Nita-Rotaru, “Key agreement for wireless networks in the presence of active adversaries,” in *Proc. IEEE Asilomar Conf. Signals, Sys. and Comp.*, 2011, pp. 792–796.
- [2] W. Halbawi, T. Ho, H. Yao, and I. Duursma, “Distributed Reed-Solomon codes for simple multiple access networks,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2014, pp. 651–655.
- [3] S. H. Dau, W. Song, Z. Dong, and C. Yuen, “Balanced sparsest generator matrices for MDS codes,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2013, pp. 1889–1893.
- [4] S. H. Dau, W. Song, and C. Yuen, “On the existence of MDS codes over small fields with constrained generator matrices,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2014, pp. 1787–1791.
- [5] ———, “On simple multiple access networks,” *IEEE J. Sel. Areas Commun. (JSAC)*, 2014, DOI: 10.1109/JSAC.2014.2384295, to appear.
- [6] M. Yan and A. Sprintson, “Algorithms for weakly secure data exchange,” in *Proc. Int. Symp. on Network Coding (NetCod)*, 2013, pp. 1–6.
- [7] M. Yan, A. Sprintson, and I. Zelenko, “Weakly secure data exchange with generalized Reed-Solomon codes,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2014, pp. 1366–1370.
- [8] T. K. Dikaliotis, T. Ho, S. Jaggi, S. Vyettrenko, H. Yao, M. Effros, J. Kliewer, and E. Erez, “Multiple-access network information-flow and correction codes,” *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 1067–1079, 2011.
- [9] H. Yamamoto, “Secret sharing system using (k, L, n) threshold scheme,” *Electronics and Communications in Japan (Part I: Communications)*, vol. 69, no. 9, pp. 46–54, 1986.
- [10] K. Bhattacharjee and K. R. Narayanan, “Weakly secure network coding,” in *Proc. 1st Workshop on Network Coding, Theory, and Application (NetCod)*, 2005.
- [11] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam: North-Holland, 1977.
- [12] S. Vladut, D. Nogin, and M. Tsfasman, *Algebraic Geometric Codes: Basic Notions*. Boston, MA, USA: American Mathematical Society, 2007.
- [13] S. H. Dau, W. Song, and C. Yuen, “On block security of regenerating codes at the MBR point for distributed storage systems,” in *Proc. Int. IEEE Symp. Inf. Theory (ISIT)*, 2014, pp. 1967–1971.
- [14] T. Ho, M. Médard, R. Koetter, D. R. Karger, M. Effros, J. Shi, and B. Leong, “A random linear network coding approach to multicast,” *IEEE Trans. Inf. Theory*, vol. 52, no. 10, 4413–4430.
- [15] P. F. Oliveira, L. Lima, T. T. V. Vinhoza, J. Barros, and M. Médard, “Coding for trusted storage in untrusted networks,” *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 6, pp. 1890–1899, 2012.

## APPENDIX

### A. Proof of Lemma 2

This lemma is a corollary of [13, Lemma 3]. We refer the reader to [13] and its extended version for a detailed and rigorous proof. Nevertheless, we provide here an informal proof that may illustrate better the intuition behind this lemma. Below we refer to the number of nonzero coordinates of a vector as its (Hamming) *weight*. It is well known in coding theory that the minimum Hamming weight of a nonzero codeword of any linear error-correcting code is equal to its minimum distance.

Suppose that every  $k - 1$  columns of the coding matrix  $\mathbf{G}$  generates an error correcting codes of minimum distance  $d \geq 2$ . Suppose that a passive adversary obtains  $\mathbf{x}\mathbf{G}[E]$ , where  $\mathbf{G}[E]$  is the  $k \times (k - 1)$  submatrix of  $\mathbf{G}$  formed by columns indexed by  $E$ , for some  $E \subset [n]$ ,  $|E| = k - 1$ . Hence, it can linearly transform these  $k - 1$  coded symbols by considering the product

$$\mathbf{x}\mathbf{G}[E]\boldsymbol{\alpha}^t = \mathbf{x}(\boldsymbol{\alpha}\mathbf{G}[E]^t)^t = \mathbf{x}\mathbf{c}^t,$$

where  $\boldsymbol{\alpha} \in \mathbb{F}_q^{k-1}$  is some coefficient vector, the superscript “ $t$ ” denotes the transpose, and  $\mathbf{c} = \boldsymbol{\alpha}\mathbf{G}[E]^t$ . Since the columns of  $\mathbf{G}[E]$  generates an error-correcting codes of minimum distance at least two,  $\mathbf{c}$ , if nonzero, has weight at least two. In other words, if  $\mathbf{c} \neq \mathbf{0}$  then it has at least two nonzero coordinates. As a result,  $\mathbf{x}\mathbf{c}^t$  is a linear combination of at least two source packets. Therefore, by linearly transforming the eavesdropped coded symbols  $\mathbf{x}\mathbf{G}[E]$ , the adversary cannot determine explicitly each source packet. As the source packets are independent and uniformly randomly distributed over  $\mathbb{F}_q$ , this is equivalent to saying that the conditional entropy of each source packet remains the same given the knowledge of  $k - 1$  coded packets. Hence, the coding scheme is weakly secure.

Conversely, if for some subset  $E \subset [n]$ ,  $|E| = k - 1$ , the columns of  $\mathbf{G}[E]$  generate a linear error-correcting code of minimum distance one, then there exists  $\boldsymbol{\alpha} \in \mathbb{F}_q^{k-1}$  such that  $\mathbf{c} = \boldsymbol{\alpha}\mathbf{G}[E]^t$  has weight one. Suppose that  $c_i \neq 0$  and  $c_j = 0$  if  $j \neq i$ . Then by post-multiplying  $\mathbf{x}\mathbf{G}[E]$  by  $\boldsymbol{\alpha}^t$ , the adversary obtains the source packet  $x_i$  explicitly. Hence, in this case the coding scheme is not weakly secure.

### B. Proof of Lemma 4

This lemma is a corollary of [3, Lemma 2-4]. Indeed, let  $\mathbf{M}$  be a  $k \times n$  binary matrix. Then [3, Lemma 2-4] conclude that  $f(\text{var}(\mathbf{M})) \neq \mathbf{0}$  if and only if

$$|\cup_{i \in I} \text{supp}(\mathbf{M}_i)| \geq n - k + |I|, \quad \forall \emptyset \neq I \subseteq [k].$$

Applying this conclusion to the  $(k - 1) \times k$  matrix  $\mathbf{P}$  in Lemma 4, the proof follows.

### C. Proof of Lemma 5

Suppose that (1) does not hold, i.e. there exists  $\emptyset \neq J \subset [n]$ ,  $|J| \leq k - 1$ , such that

$$|\cup_{j \in J} \text{supp}(\mathbf{M}[j])| \leq |J|. \quad (12)$$

We aim to show that (8) does not hold either. Indeed, from (12), let  $I \subseteq [k] \setminus \cup_{j \in J} \text{supp}(\mathbf{M}[j])$  such that  $|I| = k - |J|$ . Because  $1 \leq |J| \leq k - 1$ , we deduce that  $\emptyset \neq I \subsetneq [k]$ . Moreover, due to (12) and our assumption that  $I \subseteq [k] \setminus \cup_{j \in J} \text{supp}(\mathbf{M}[j])$ , we conclude that

$$|\cup_{i \in I} \text{supp}(\mathbf{M}_i)| \leq n - |J| = n - k + |I|.$$

Hence, (8) is violated.

Conversely, we need to show that if (8) does not hold then neither does (1). The proof is completely similar and therefore is omitted.

### D. Proof of Theorem 2

We can prove this theorem by modifying the proof of [4, Theorem 2] accordingly. Both proofs follow the same idea of a well-known proof of Hall's marriage theorem: repeatedly removing the edges of the bipartite graph until the graph becomes sparsest yet still satisfies the Hall's condition. To simplify the notation, for a set  $I \subseteq [k]$  we use  $R_I$  to denote the union  $\cup_{i \in I} R_i$ .

Suppose that the sets  $R_i$  satisfy (10). We keep removing the elements of these sets while maintaining the Weak Security Condition (10). Assume that at some point, the removal of any element in any set  $R_i$  would make them violate (10). We prove that now the sets  $R_i$  have cardinality precisely  $n - k + 2$ , which concludes the first part of the theorem.

Suppose, for contradiction, that there exists  $r \in [k]$  such that  $|R_r| \geq n - k + 3$ . Take  $a$  and  $b$  in  $R_r$ ,  $a \neq b$ . For all  $i \in [k]$ , let

$$R_i^a = \begin{cases} R_i \setminus \{a\}, & \text{if } i = r, \\ R_i, & \text{otherwise,} \end{cases} \quad (13)$$

$$R_i^b = \begin{cases} R_i \setminus \{b\}, & \text{if } i = r, \\ R_i, & \text{otherwise.} \end{cases} \quad (14)$$

According to our assumption, both of the two collections of sets  $\{R_i^a\}_{i \in [k]}$  and  $\{R_i^b\}_{i \in [k]}$  violate (10). Therefore, there exist two nonempty subsets  $A \subseteq [k]$  and  $B \subseteq [k]$ ,  $r \notin A \cup B$ , such that

$$|R_{A \cup \{r\}}^a| < n - k + |A| + 2, \quad (15)$$

$$|R_{B \cup \{r\}}^b| < n - k + |B| + 2. \quad (16)$$

Since  $r \notin A$ , by (13) we have

$$|R_{A \cup \{r\}}^a| \geq |R_A^a| = |R_A| \geq n - k + |A| + 1. \quad (17)$$

Similarly, since  $r \notin B$ , by (14) we have

$$|R_{B \cup \{r\}}^b| \geq |R_B^b| = |R_B| \geq n - k + |B| + 1. \quad (18)$$

From (15) and (17) we deduce that

$$|R_{A \cup \{r\}}^a| = |R_A^a| = |R_A| = n - k + |A| + 1. \quad (19)$$

Similarly, from (16) and (18) we have

$$|R_{B \cup \{r\}}^b| = |R_B^b| = |R_B| = n - k + |B| + 1. \quad (20)$$

Therefore,

$$R_{A \cup \{r\}}^a \cap R_{B \cup \{r\}}^b = R_A \cap R_B. \quad (21)$$

Moreover, as  $a \in R_{B \cup \{r\}}^b$  and  $b \in R_{A \cup \{r\}}^a$ , we deduce that

$$R_{A \cup \{r\}}^a \cup R_{B \cup \{r\}}^b = R_{A \cup B \cup \{r\}}. \quad (22)$$

From (19) and (20) we have

$$\begin{aligned} & 2(n - k) + |A| + |B| + 2 \\ &= |R_{A \cup \{r\}}^a| + |R_{B \cup \{r\}}^b| \\ &= |R_{A \cup \{r\}}^a \cup R_{B \cup \{r\}}^b| + |R_{A \cup \{r\}}^a \cap R_{B \cup \{r\}}^b| \\ &= |R_{A \cup B \cup \{r\}}| + |R_A \cap R_B|, \end{aligned} \quad (23)$$

where the last transition is due to (21) and (22). We further evaluate the two terms of the last sum in (23) as follows. The first term

$$\begin{aligned} |R_{A \cup B \cup \{r\}}| &\geq n - k + |A \cup B \cup \{r\}| + 1 \\ &= n - k + |A \cup B| + 2. \end{aligned} \quad (24)$$

The second term

$$|R_A \cap R_B| \geq n - k + |A \cap B| + 1, \quad (25)$$

which can be explained below.

- If  $A \cap B \neq \emptyset$ , then by applying (10) to  $A \cap B$  we obtain

$$|R_A \cap R_B| \geq |R_{A \cap B}| \geq n - k + |A \cap B| + 1.$$

- If  $A \cap B = \emptyset$ , then  $n - k + |A \cap B| + 1 = n - k + 1$ . We have

$$R_{A \cup \{r\}}^a = R_A^a \cup R_r^a = R_A \cup (R_r \setminus \{a\}). \quad (26)$$

By (19),  $R_{A \cup \{r\}}^a = R_A$ . Combining this with (26) we deduce that

$$R_r \setminus \{a\} \subseteq R_A. \quad (27)$$

Similarly,

$$R_r \setminus \{b\} \subseteq R_B. \quad (28)$$

From (27) and (28) we have

$$|R_A \cap R_B| \geq |R_r \setminus \{a, b\}| \geq (n - k + 3) - 2 = n - k + 1,$$

which proves that (25) is correct when  $A \cap B = \emptyset$ .

Finally, from (23), (24), and (25) we deduce that

$$\begin{aligned} & 2(n - k) + |A| + |B| + 2 \\ & \geq (n - k + |A \cup B| + 2) + (n - k + |A \cap B| + 1) \\ & = 2(n - k) + |A| + |B| + 3, \end{aligned}$$

which produces a contradiction.

The proof of the first part of this theorem also provides a polynomial time algorithm to find subsets of  $R_i$ 's that all have cardinality  $n - k + 2$  yet still maintain the Weak Security Condition (10). Indeed, we keep removing the elements of the subsets  $R_i$  in the following way. If there exists  $r \in [k]$  such that  $|R_r| \geq n - k + 3$ , then as we just prove, for  $a, b \in R_r$ , it is impossible that removing  $a$  or  $b$  from  $R_r$  both render the Weak Security Condition violated. Therefore,

we can either remove  $a$  or  $b$  while still maintaining the Weak Security Condition. Note that by Proposition 1, the Weak Security Condition can be verified in polynomial time in  $k$  and  $n$ . Therefore, this algorithm terminates in polynomial time in  $k$  and  $n$  and produces subsets  $R'_i$ 's of the original sets  $R_i$ 's that satisfy the stated requirement in the theorem. Note that by setting  $I = \{i\}$ , the Weak Security Condition (10) implies that  $|R'_i| \geq n - k + 2$ . Hence, those  $R'_i$ 's form a sparsest  $(n, k)$ -SMAN that still supports a weakly secure MDS coding scheme.